A Hilbert Space Central Limit Theorem for Geometrically Ergodic Markov Chains☆

John Stachurski

Research School of Economics, Australian National University

Abstract

This note proves a simple but useful central limit theorem for Hilbert space valued functions of geometrically ergodic Markov chains on general state spaces. The theorem is valid for chains starting at an arbitrary point in the state space.

Keywords: Markov chains, Hilbert space, Central limit theorem

1. Introduction

Let $(X_t)_{t \geq 1}$ be a geometrically ergodic Markov chain on state space $\mathbb{X}$ (full definitions follow) and let π be the unique stationary distribution. It is well-known (see, for example, [5] or [8, chapter 17]) that if T is a measurable function from the state space $\mathbb{X}$ to $\mathbb{R}$ satisfying a suitable second moment condition, then $n^{-1/2} \sum_{t=1}^{n} [T(X_t) - \int T d\pi]$ converge in law to a centered Gaussian distribution on $\mathbb{R}$. Using the Cramer-Wold device, the same result can be extended without technical difficulties to the case where T takes values in $\mathbb{R}^n$.[1]

In this paper we provide an analogous CLT result for the case where T takes values in is a separable Hilbert space. The aim is not to provide a particularly general Hilbert central limit theorem for dependent variables, but rather to provide a set of conditions that are straightforward to check in applications. The proof of our result is based on the dependent variable Hilbert CLT of Merlevède *et al.* [7].

2. Set Up

Let $(\Omega, \mathscr{F}, \mathbb{P})$ denote an arbitrary probability space on which all random variables are supported. As usual, if $(E, \mathscr{B})$ is any measurable space, then an E-valued random variable X is a measurable map from $(\Omega, \mathscr{F})$ to $(E, \mathscr{B})$. We use the symbol $\mathcal{L}X$ to denote its law (i.e., $\mathcal{L}X = \mathbb{P} \circ X^{-1}$). In what follows, if E has a topology then, the σ-algebra $\mathscr{B}$ is always taken to be the Borel sets. Unless otherwise stated, measurability of functions refers to Borel measurability. If μ is a measure on $(E, \mathscr{B})$ and h is a real-valued measurable function on E, then $\mu(h)$ denotes $\int h d\mu$ whenever the latter is defined. If E is a topological space and $(\mu_n)_{n \geq 0}$ are probabilities (i.e., Borel probability measures) on E, then $\mu_n \to \mu_0$ in distribution if $\mu_n(h) \to \mu_0(h)$ in $\mathbb{R}$ for every continuous bounded $h \colon E \to \mathbb{R}$. The sequence $(\mu_n)_{n \geq 1}$ is called *tight* if for all $\varepsilon > 0$ there is a compact $K \subset E$ with $\sup_{n \geq 1} \mu_n(E \setminus K) \leq \varepsilon$.

Below we consider a stochastic process taking values in a separable Hilbert space $\mathcal{H}$. Let $\| \cdot \|$ denote the norm on $\mathcal{H}$, and $\langle h, g \rangle$ the inner product of h and g. If Y is an $\mathcal{H}$-valued random variable with $\mathbb{E}\|Y\| < \infty$, then, by the Riesz representation theorem, there exists a unique element $\mathcal{E}Y$ of $\mathcal{H}$ such that $\mathbb{E}\langle h, Y \rangle =$

☆Financial support from Australian Research Council grant DP120100321 is gratefully acknowledged.
Email address: `john.stachurski@anu.edu.au` (John Stachurski)

[1]For applications of this CLT focusing on Markov chain Monte Carlo, see the surveys of Roberts and Rosenthal [10] and Jones [5]. For an application in economics see [9]. For other applications see [8].

$\langle h, \mathcal{E}Y \rangle$ for all $h \in \mathcal{H}$. The vector $\mathcal{E}Y$ is called the *expectation* (or Pettis integral) of Y. For any $\mathcal{H}$-valued random variable Y with $\mathbb{E}\|Y\|^2 < \infty$ and $\mathcal{E}Y = 0$, the *covariance operator* $C\colon \mathcal{H} \to \mathcal{H}$ of Y is defined by $\langle g, Ch \rangle = \mathbb{E}\langle g, Y \rangle \langle h, Y \rangle$ for all $g, h \in \mathcal{H}$. A random variable V taking values in $\mathcal{H}$ is called *Gaussian* if $\langle h, V \rangle$ is Gaussian on $\mathbb{R}$ for each $h \in \mathcal{H}$. To simplify the presentation, in what follows we regard degenerate random variables on $\mathbb{R}$ as Gaussians with zero variance.[2]

3. Main Result

Let $(\mathbb{X}, \mathscr{X})$ be a measure space, and let P be a stochastic kernel on $\mathbb{X}$. In particular, $P(x, dy)$ is a probability measure on $(\mathbb{X}, \mathscr{X})$ for each $x \in \mathbb{X}$, and $x \mapsto P(x, B)$ is measurable for every $B \in \mathscr{X}$. In what follows, we use the standard notation

$$(\psi P)(B) := \int P(x, B)\psi(dx) \quad \text{and} \quad (Pf)(x) := \int f(y) P(x, dy).$$

Here ψ is a probability measure on $(\mathbb{X}, \mathscr{X})$ and $f\colon \mathbb{X} \to \mathbb{R}$ is a measurable function such that the integral is defined. Let P^t denote the t-th iterate of either one of these operators. A probability π on $(\mathbb{X}, \mathscr{X})$ is called stationary for P if $\pi P = \pi$.

Let $\| \cdot \|_{TV}$ be the total variation norm over the space of finite signed measures on $(\mathbb{X}, \mathscr{X})$. We assume throughout that P is geometrically ergodic, which is to say that (i) P has a unique stationary distribution π, (ii) $\|\psi P^t - \varphi P^t\|_{TV} \to 0$ as $t \to \infty$ for any probabilities ψ and φ on $(\mathbb{X}, \mathscr{X})$, and (iii) there exists a measurable function $V\colon \mathbb{X} \mapsto [0, \infty)$ and constants $R \in \mathbb{R}_+$ and $\alpha \in [0, 1)$ such that

$$\int V d\pi < \infty \quad \text{and} \quad \sup_{B \in \mathscr{X}} \left| P^t(x, B) - \pi(B) \right| \leq \alpha^t R V(x) \quad \text{for all } x \in \mathbb{X},\ t \in \mathbb{N} \tag{1}$$

Sufficient conditions for geometric ergodicity are discussed in many sources. See, for example, [8] and [4]. See also [6, Theorem 21.12] for a range of conditions equivalent to (ii).

Letting ψ be a probability measure on $\mathbb{X}$, we call an $\mathbb{X}$-valued stochastic process $(X_t)_{t \geq 1}$ *Markov-*(P, ψ) if X_1 is drawn from ψ and P is the transition probability function for $(X_t)_{t \geq 1}$. More formally, this means that

$$\mathbb{E}[h(X_{t+k}) \mid \mathscr{F}_t] = P^k h(X_t) \tag{2}$$

almost surely for any $t, k \in \mathbb{N}$ and any bounded measurable $h\colon \mathbb{X} \to \mathbb{R}$, and, in addition, $\mathcal{L}X_1 = \psi$. Here $\mathscr{F}_t$ is the σ-algebra generated by $(X_1, \ldots, X_t)$, and $\mathbb{E}[\cdot \mid \mathscr{F}_t]$ is conditional expectation with respect to $\mathscr{F}_t$. Existence of at least one such a sequence $(X_t)_{t \geq 1}$ follows from a well-known theorem of Ionescu-Tulcea (see, e.g., [11, theorem II.9.2]). If ψ is a Dirac probability measure concentrated at a single point x, then we call $(X_t)_{t \geq 1}$ *Markov-*(P, x). If $(X_t)_{t \geq 1}$ is Markov-(P, π), then $(X_t)_{t \geq 1}$ is stationary, and $\mathcal{L}X_t = \pi$ for all t (see, e.g., [8, chapter 3]).

Our main result concerns sequences of the form $[T_0(X_t)]_{t \geq 1}$, where T_0 is a measurable map from $\mathbb{X}$ into a separable Hilbert space $\mathcal{H}$. On T_0 we impose the following assumption:

Assumption 3.1. There exists nonnegative constants m_0, m_1 and $\gamma < 1$ such that

$$\|T_0(x)\|^2 \leq m_0 + m_1 V(x)^\gamma \quad \text{for all} \quad x \in \mathbb{X}.$$

The following lemma assures us that if $\mathcal{L}X = \pi$, then $\mathcal{E}T_0(X)$ exists.

Lemma 3.1. *If $\mathcal{L}X = \pi$ and assumption 3.1 holds, then $\mathbb{E}\|T_0(X)\| < \infty$.*

[2]For more details on Hilbert-space valued stochastic processes, see, for example, [1].

Proof. Assume the conditions of the lemma. It suffices to show that $\mathbb{E}\|T_0(X)\|^2 < \infty$. Applying assumption 3.1 and Jensen's inequality, we have

$$\mathbb{E}\|T_0(X)\|^2 \leq m_0 + m_1 \mathbb{E}[V(X)^\gamma] \leq m_0 + m_1 [\mathbb{E}V(X)]^\gamma$$

The final expression is finite by the left-hand side of (1). □

We need two final definitions. Let $(X_t)_{t\geq 1}$ be Markov-(P, π). By lemma 3.1, $\mathcal{E}T_0(X_1)$ exists in $\mathcal{H}$. Define $T\colon \mathbb{X} \to \mathcal{H}$ be the map

$$T(x) = T_0(x) - \mathcal{E}T_0(X_1) \qquad (x \in \mathbb{X}),$$

and let C be the covariance operator defined by

$$\langle g, Ch \rangle = \mathbb{E}\langle g, T(X_1)\rangle \langle h, T(X_1)\rangle + \sum_{t\geq 2} \mathbb{E}\langle g, T(X_1)\rangle \langle h, T(X_t)\rangle + \sum_{t \geq 2} \mathbb{E}\langle h, T(X_1)\rangle \langle g, T(X_t)\rangle. \tag{3}$$

for $g, h \in \mathcal{H}$. We can now state our main result:

Theorem 3.1. *Let assumption 3.1 hold. If $x \in \mathbb{X}$ and $(X_t)_{t\geq 1}$ is Markov-(P, x), then*

$$\mathcal{L}\left[n^{-1/2} \sum_{t=1}^{n} T(X_t) \right] \to N(0, C) \qquad (n \to \infty). \tag{4}$$

Here $N(0, C)$ represents the distribution of an $\mathcal{H}$-valued Gaussian random variable with expectation equal to the origin of $\mathcal{H}$ and covariance operator C.

3.1. Example

Before turning to the proof of theorem 3.1, we present a simple illustration. Let μ be any probability measure on $(\mathbb{R}, \mathscr{B})$, and consider the separable Hilbert space $L_2 := L_2(\mathbb{R}, \mathscr{B}, \mu)$. Let P be a geometrically ergodic stochastic kernel on $\mathbb{R}$, and let F be the cumulative distribution function of its stationary distribution. In many cases, no closed form expression for F is available. Suppose that we wish to compute it by simulation. A natural technique is to pick any $x \in \mathbb{R}$, simulate a Markov-(P, x) process $(X_t)_{t\geq 1}$, and evaluate the empirical cumulative distribution function $F_n(y) := \frac{1}{n}\sum_{t=1}^{n} \mathbb{1}\{X_t \leq y\}$. Let us investigate the error $F_n - F$, measured in L_2 norm. Define $T_0(x)$ to be the function $y \mapsto \mathbb{1}\{x \leq y\}$. We then have $\|T_0(x)\|^2 = \int \mathbb{1}\{x \leq y\}^2 \mu(dy) = \mu([x, \infty)) \leq 1$. Taking $m_0 = 1$ and $m_1 = 0$, we see that assumption 3.1 is alway satisfied. Moreover, a straightforward application of Fubini's theorem shows that if $\mathcal{L}X_1 = F$, then $\mathcal{E}T_0(X_1) = F$. As a result, setting $T := T_0 - F$, theorem 3.1 gives

$$\sqrt{n}(F_n - F) = \sqrt{n}\left\{ \frac{1}{n} \sum_{t=1}^{n} T_0(X_t) - F \right\} = n^{-1/2} \sum_{t=1}^{n} T(X_t) \to N(0, C)$$

where C is defined by (3). As a corollary, continuity of the norm now implies that $\|F_n - F\| = O_P(n^{-1/2})$.

4. Proof of theorem 3.1

Our first lemma shows that, given our ergodicity assumptions on P, we can restrict attention to the case where $\mathcal{L}X_1 = \pi$ when proving (4).

Lemma 4.1. *Let $(X_t)_{t\geq 1}$ and $(X'_t)_{t\geq 1}$ be two P-Markov chains, where $\mathcal{L}X_1 = \pi$ and $X'_1 = x \in \mathbb{X}$. For any Borel probability measure ν on $L_2(\mu)$,*

$$\mathcal{L}\left[n^{-1/2} \sum_{t=1}^{n} T(X_t) \right] \to \nu \quad \textit{implies} \quad \mathcal{L}\left[n^{-1/2} \sum_{t=1}^{n} T(X'_t) \right] \to \nu$$

Proof. Given our assumption of geometric ergodicity (and hence ergodicity), it is well known (see Lindvall, [6, Theorem 21.12]) that one can construct P-Markov processes $(X_t)_{t\geq 1}$ and $(X'_t)_{t\geq 1}$ on a common probability space $(\Omega, \mathscr{F}, \mathbb{P})$ such that

$$\tau := \inf\{t \in \mathbb{N} : X_t = X'_t\}$$

is finite almost surely, and $X_t = X'_t$ for all $t \geq \tau$. Let $S_n := \sum_{t=1}^n T(X_t)$ and $S'_n := \sum_{t=1}^n T(X'_t)$, and assume as in the statement of the lemma that $n^{-1/2}S_n \to \nu$. To prove that $n^{-1/2}S'_n \to \nu$ it suffices to show that the (norm) distance between $n^{-1/2}S'_n$ and $n^{-1/2}S_n$ converges to zero in probability (cf., e.g., Dudley, [3, Lemma 11.9.4]). Fixing $\varepsilon > 0$, we need to show that

$$\mathbb{P}\{\|n^{-1/2}S'_n - n^{-1/2}S_n\| > \varepsilon\} \to 0 \quad (n \to \infty) \tag{5}$$

Clearly

$$\{\|n^{-1/2}S'_n - n^{-1/2}S_n\| > \varepsilon\} \subset \left\{\sum_{t=1}^n \|T(X'_t) - T(X_t)\| > n^{1/2}\varepsilon\right\}$$

Fix $k \in \mathbb{N}$, and partition the last set over $\{\tau \leq k\}$ and $\{\tau > k\}$ to obtain the disjoint sets

$$\left\{\sum_{t=1}^n \|T(X'_t) - T(X_t)\| > n^{1/2}\varepsilon\right\} \cap \{\tau \leq k\} \subset \left\{\sum_{t=1}^k \|T(X'_t) - T(X_t)\| > n^{1/2}\varepsilon\right\}$$

and

$$\left\{\sum_{t=1}^n \|T(X'_t) - T(X_t)\| > n^{1/2}\varepsilon\right\} \cap \{\tau > k\} \subset \{\tau > k\}$$

Together, these lead to the bound

$$\{\|n^{-1/2}S'_n - n^{-1/2}S_n\| > \varepsilon\} \subset \left\{\sum_{t=1}^k \|T(X'_t) - T(X_t)\| > n^{1/2}\varepsilon\right\} \cup \{\tau > k\}$$

$$\therefore \quad \mathbb{P}\{\|n^{-1/2}S'_n - n^{-1/2}S_n\| > \varepsilon\} \leq \mathbb{P}\left\{\sum_{t=1}^k \|T(X'_t) - T(X_t)\| > n^{1/2}\varepsilon\right\} + \mathbb{P}\{\tau > k\}$$

For any fixed k, we have

$$\lim_{n\to\infty} \mathbb{P}\left\{\sum_{t=1}^k \|T(X'_t) - T(X_t)\| > n^{1/2}\varepsilon\right\} = 0 \tag{6}$$

Hence

$$\limsup_{n\to\infty} \mathbb{P}\{\|n^{-1/2}S'_n - n^{-1/2}S_n\| > \varepsilon\} \leq \mathbb{P}\{\tau > k\}, \quad \forall k \in \mathbb{N}$$

Since $\mathbb{P}\{\tau < \infty\} = 1$ taking $k \to \infty$ yields (5). ☐

In view of Lemma 4.1, we can continue the proof of (4) while considering only the case $\mathcal{L}X_1 = \pi$. In this case $(T(X_t))$ is a centered strict sense stationary stochastic processes in $\mathcal{H}$, and we can apply the stationary Hilbert CLT in Merlevède *et al.* [7, Theorem 4, Corollary 1]. From the latter we obtain the following result: Let $\xi_t := T(X_t)$ for all t. Define the corresponding mixing coefficients by

$$\alpha(t) := \sup |\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)|$$

where the supremum is over all $A \in \sigma(\xi_1)$ and $B \in \sigma(\xi_{t+1})$. In this setting, the convergence in (4) will be valid whenever there exists a constant $\delta > 0$ such that

$$\mathbb{E}\|\xi_t\|^{2+\delta} < \infty \quad \text{and} \quad \sum_{t=1}^\infty t^{2/\delta}\alpha(t) < \infty \tag{7}$$

(The definition of the mixing coefficient used here is slightly different to the one used in Merlevède *et al.* [7, Definition 1]. However, in the Markov case it is well-known that the two are equivalent. See, for example, Bradley [2, Section 3].)

We establish first the finite expectation on the left-hand side of (7). Let m_0, m_1, γ and V be the constants and function in assumption 3.1. Let $r := \|\mathcal{E}T(X_1)\|^2$. Evidently

$$\|T(x)\|^{2/\gamma} = \|T_0(x) - \mathcal{E}T(X_1)\|^{2/\gamma} \leq \left[2\|T_0(x)\|^2 + 2r\right]^{1/\gamma}$$

From this bound, assumption 3.1 and Jensen's inequality, we obtain

$$\|T(x)\|^{2/\gamma} \leq [2m_0 + 2m_1 V(x)^\gamma + 2r]^{1/\gamma} \leq \frac{1}{3}\left\{[6m_0]^{1/\gamma} + [6m_1 V(x)^\gamma]^{1/\gamma} + [6r]^{1/\gamma}\right\}$$

In other words, there exist finite constants c_1 and c_2 such that

$$\|\xi_t\|^{2/\gamma} := \|T(X_t)\|^{2/\gamma} \leq c_1 V(X_t) + c_2$$

holds pointwise on Ω. Let $\delta := 2(1-\gamma)/\gamma$, so that $2/\gamma = 2 + \delta$. Taking expectations and applying the first expression in (1) gives $\mathbb{E}\|\xi_t\|^{2+\delta} < \infty$ as required.

The last step of the proof of Theorem 3.1 is to verify the finiteness of the sum on the right-hand side of (7). An elementary argument shows the following ordering of σ-algebras:

$$\sigma(\xi_j) = \sigma(T(X_j)) \subset \sigma(X_j), \quad \forall j$$

As a result, we have

$$\alpha(t) := \sup_{\substack{A\in\sigma(\xi_1) \\ B\in\sigma(\xi_{t+1})}} |\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)| \leq \sup_{\substack{A\in\sigma(X_1) \\ B\in\sigma(X_{t+1})}} |\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)|$$

The right-hand side gives the strong mixing coefficients for (X_t), which, in the geometrically ergodic case, are known to be $O(\lambda^t)$ for the constant λ in (1). (See, for example, Jones [5, p. 304].) As a consequence, we have $\alpha(t) = O(\lambda^t)$, and hence $\sum_{t=1}^{\infty} t^{2/\delta}\alpha(t)$ will be finite if $\sum_{t=1}^{\infty} t^{2/\delta}\lambda^t$ is finite. Since $\lambda < 1$, this last sum is clearly finite. This completes the proof of Theorem 3.1.

References

[1] Bosq, D., 2000. Linear Processes in Function Space, Springer-Verlag.

[2] Bradley, R. C. (2005): "Basic properties of strong mixing conditions: A survey and some open questions," *Probability Surveys*, 2, 107–144.

[3] Dudley, Richard M. (2002): *Real Analysis and Probability*, Cambridge Studies in Advanced Mathematics No. 74, Cambridge University Press.

[4] Hairer, M. and J. C. Mattingly (2011): "Yet another look at Harris' ergodic theorem for Markov chains," in *Seminar on Stochastic Analysis, Random Fields and Applications VI* (R. C. Dalang, M. Dozzi and F. Russo, ed.) Springer Basel.

[5] Jones, G.L., 2004. On the Markov chain central limit theorem, Probab. Surv., 1, 299–320.

[6] Lindvall, T. (2002): *Lectures on the Coupling Method*, Dover Publications, Mineola N.Y.

[7] Merlevède, F., M. Peligrad and S. Utev (1997): "Sharp conditions for the CLT of Linear Processes in a Hilbert Space," *Journal of Theoretical Probability*, 10 (3), 681-693.

[8] S. Meyn, Tweedie, R.L., 2009. Markov Chains and Stochastic Stability, 2nd Edition, Cambridge University Press, Cambridge.

[9] Nishimura, K., Stachurski, J., 2005. Stability of stochastic optimal growth models: A New Approach, J. Econ. Theory, 122 (1), 100–118.

[10] Roberts, G.O., Rosenthal, J.S., 2004. General state Markov chains and MCMC algorithms, Probab. Surv., 1, 20–71.

[11] Shiryaev, A.N., 1996. Probability, Springer-Verlag, New York.